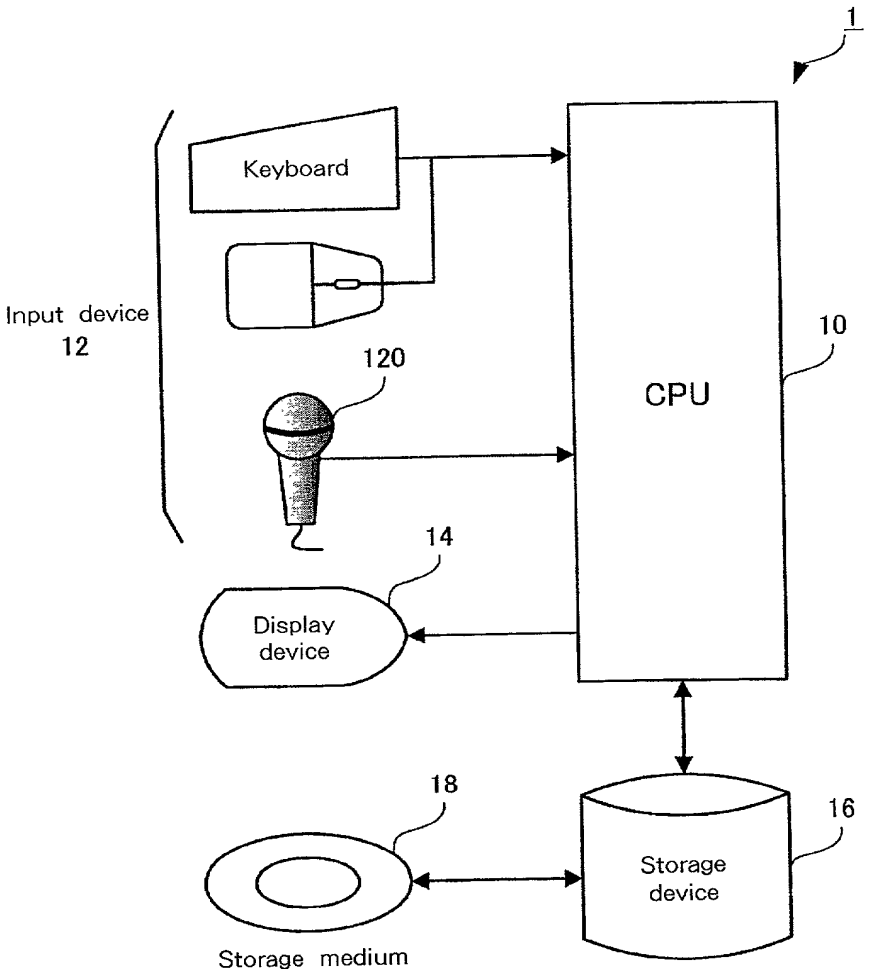


Fig. 1

EL649719607US

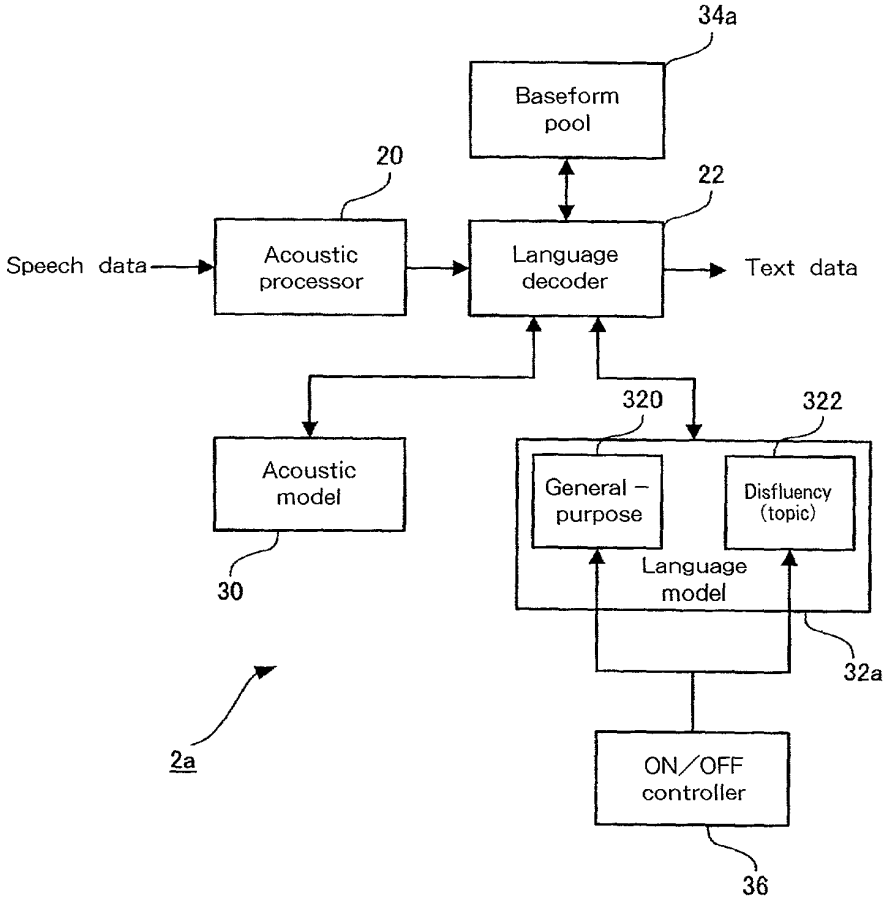


Fig. 2

(a) Case wherein a disfluency language model is OFF
"..., koregaehonhatsumeinoyouten ..."
→ "..., kore ga e – hon hatsu – mei no you – ten"

(b) Case wherein a disfluency language model is ON
"..., koregaehonhatsumeinoyouten ..."
→ "..., kore ga <er> hon – hatsumei no you – ten ..."

Fig. 3

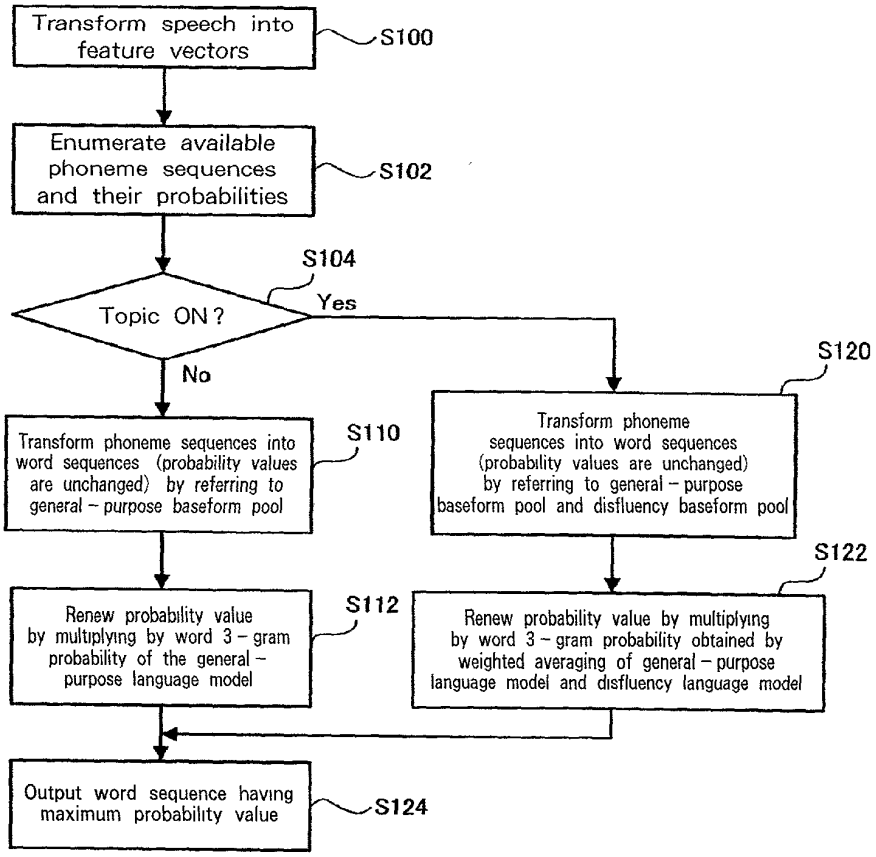

Transform speech into feature vectors
S100
Enumerate available phoneme sequences and their probabilities
S102
S104
Topic ON?
Yes
No
S120
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general – purpose baseform pool
S110
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general – purpose baseform pool and disfluency baseform pool
S122
Renew probability value by multiplying by word 3 – gram probability of the general – purpose language model
S112
Renew probability value by multiplying by word 3 – gram probability obtained by weighted averaging of general – purpose language model and disfluency language model
Output word sequence having maximum probability value
S124

Fig. 4

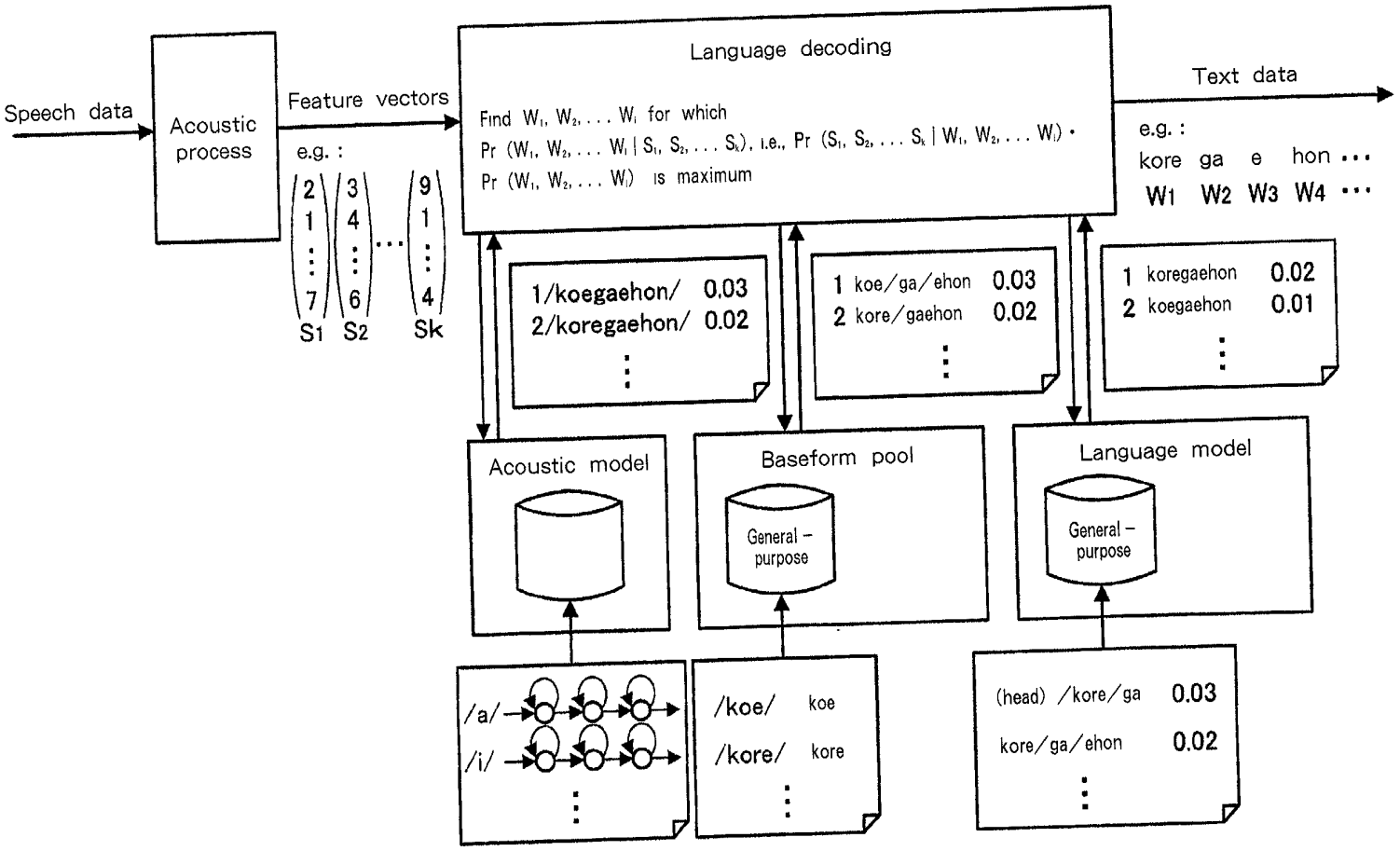

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wi for which
Pr (W1, W2, ... Wi | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wi) ·
Pr (W1, W2, ... Wi) is maximum
Text data
e.g. :
kore ga e hon ...
W1 W2 W3 W4 ...
1/koegaehon/ 0.03
2/koregaehon/ 0.02
1 koe/ga/ehon 0.03
2 kore/gaehon 0.02
1 koregaehon 0.02
2 koegaehon 0.01
Acoustic model
Baseform pool
Language model
General – purpose
General – purpose
/a/
/i/
/koe/ koe
/kore/ kore
(head) /kore/ga 0.03
kore/ga/ehon 0.02

Fig. 5

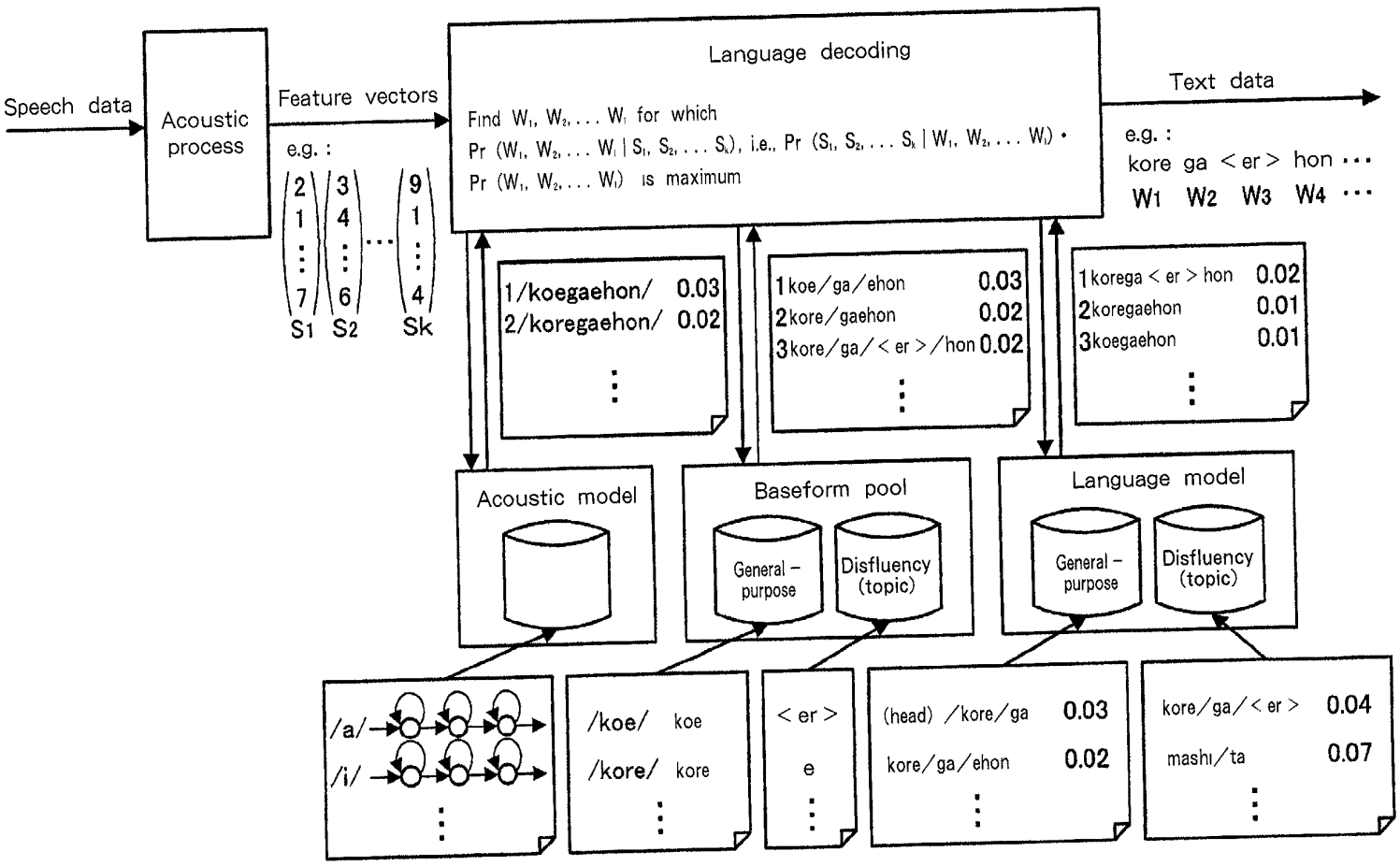

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wi for which
Pr (W1, W2, ... Wi | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wi) •
Pr (W1, W2, ... Wi) is maximum
Text data
e.g. :
kore ga < er > hon ...
W1 W2 W3 W4 ...
1/koegaehon/ 0.03
2/koregaehon/ 0.02
1koe/ga/ehon 0.03
2kore/gaehon 0.02
3kore/ga/< er >/hon 0.02
1korega < er > hon 0.02
2koregaehon 0.01
3koegaehon 0.01
Acoustic model
Baseform pool
General – purpose
Disfluency (topic)
Language model
General – purpose
Disfluency (topic)
/a/
/i/
/koe/ koe
/kore/ kore
< er >
e
(head) /kore/ga 0.03
kore/ga/ehon 0.02
kore/ga/< er > 0.04
mashi/ta 0.07

Fig. 6

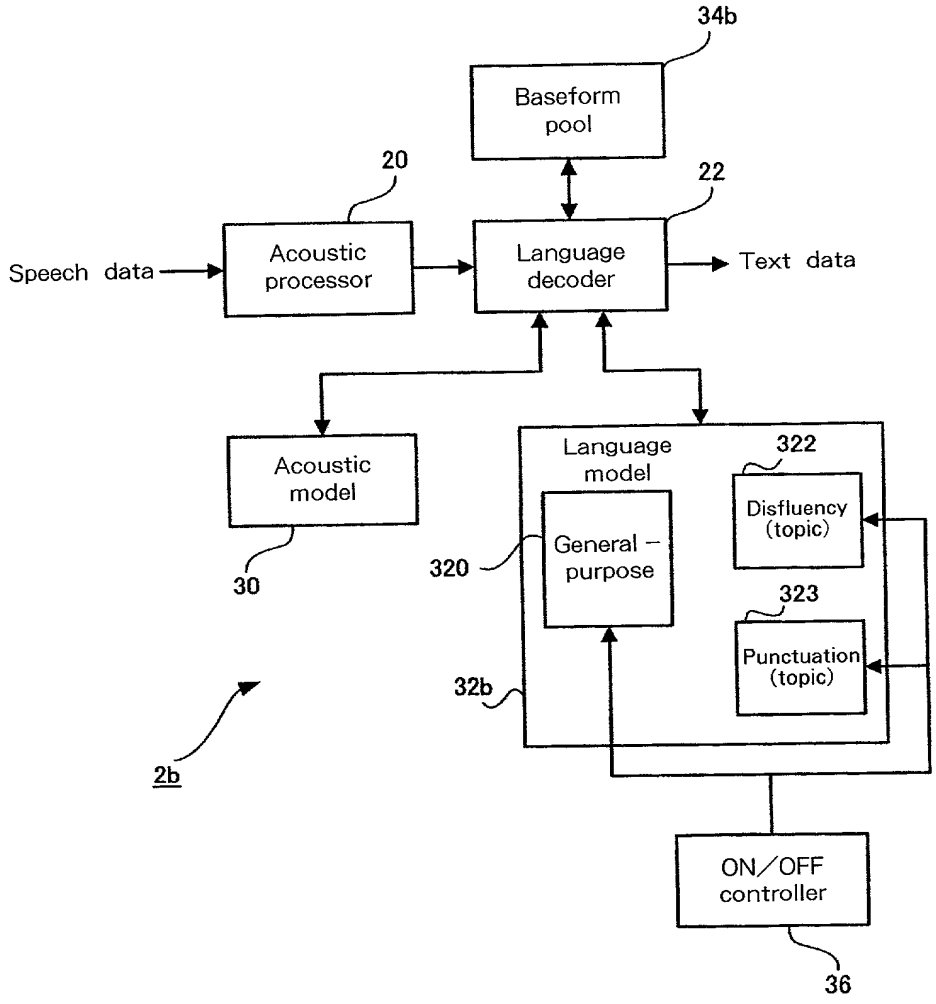

34b
Baseform pool
20
22
Speech data
Acoustic processor
Language decoder
Text data
Acoustic model
30
Language model
322
Disfluency (topic)
320
General – purpose
323
Punctuation (topic)
32b
2b
ON/OFF controller
36

Fig. 7

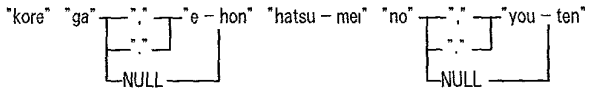
(a) Case wherein a disfluency language model is OFF and a punctuation language model is OFF

"..., korega (pause) (ten) ehonhatsumeino (pause) youten ..."

→ "..., kore ga, e – hon – hatsu – mei no you – ten ..."

(b) Case wherein a disfluency language model is OFF and a punctuation language model is ON

"..., korega (pause) ehonhatsumeino (pause) youten ..."



→ "..., kore ga, e – hon – hatsu – mei no you – ten ..."

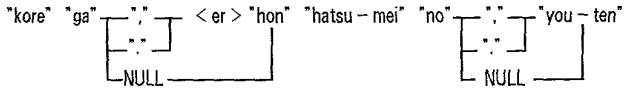
(c) Case wherein a disfluency language model is ON and a punctuation language model is OFF

"..., korega (pause) (ten) ehonhatsumeino (pause) youten ..."

→ "..., kore ga, < er > hon – hatsu – mei no you – ten ..."

(d) Case wherein a disfluency language model is OFF and a punctuation language model is ON

"..., korega (pause) ehonhatsumeino (pause) youten ..."



→ "..., kore ga, < er > hon – hatsu – mei no you – ten ..."

Fig. 8

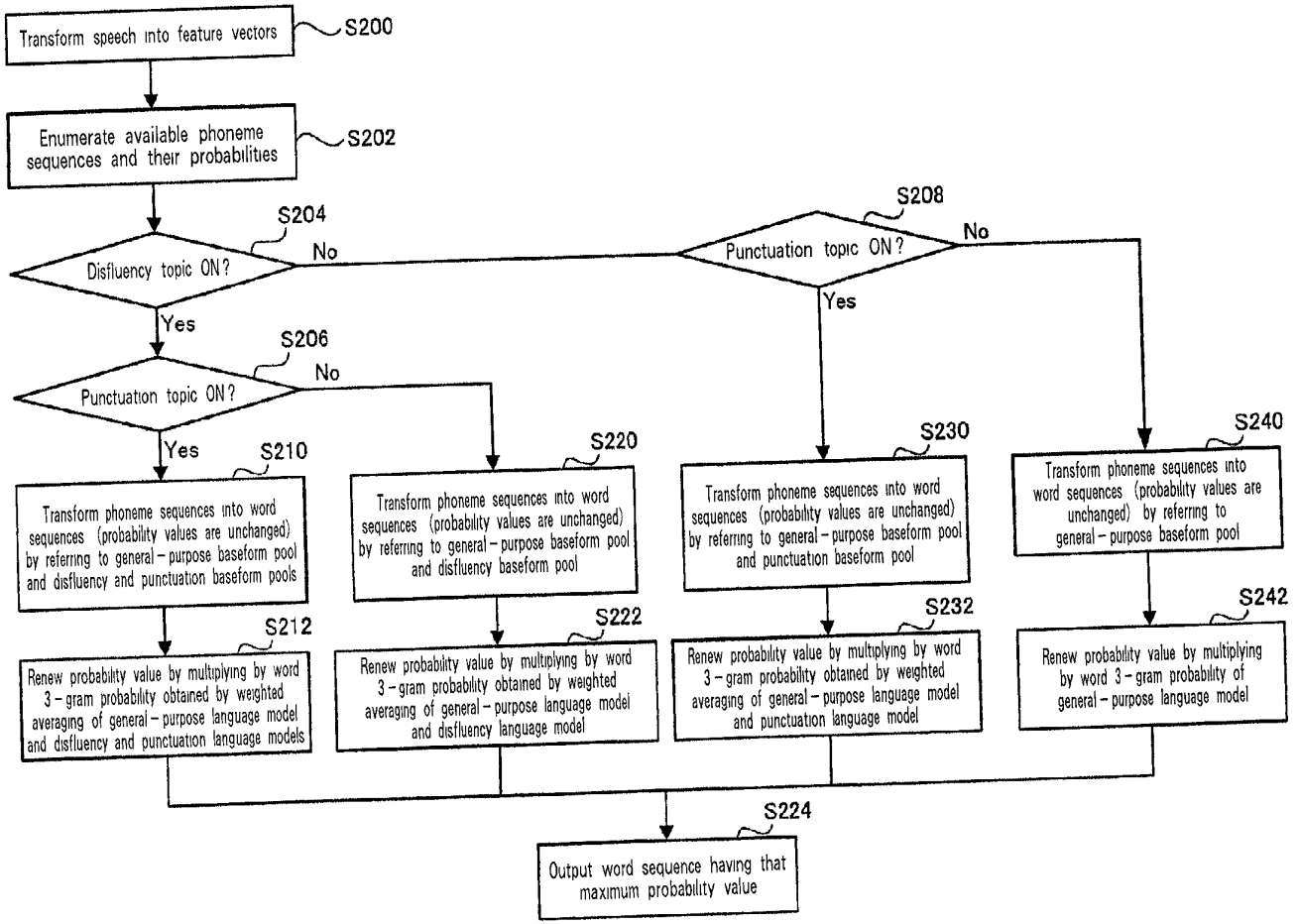

Transform speech into feature vectors
S200
Enumerate available phoneme sequences and their probabilities
S202
Disfluency topic ON?
S204
No
Yes
Punctuation topic ON?
S206
No
Yes
Punctuation topic ON?
S208
No
Yes
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general-purpose baseform pool and disfluency and punctuation baseform pools
S210
Renew probability value by multiplying by word 3-gram probability obtained by weighted averaging of general-purpose language model and disfluency and punctuation language models
S212
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general-purpose baseform pool and disfluency baseform pool
S220
Renew probability value by multiplying by word 3-gram probability obtained by weighted averaging of general-purpose language model and disfluency language model
S222
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general-purpose baseform pool and punctuation baseform pool
S230
Renew probability value by multiplying by word 3-gram probability obtained by weighted averaging of general-purpose language model and punctuation language model
S232
Transform phoneme sequences into word sequences (probability values are unchanged) by referring to general-purpose baseform pool
S240
Renew probability value by multiplying by word 3-gram probability of general-purpose language model
S242
Output word sequence having that maximum probability value
S224

Fig. 9

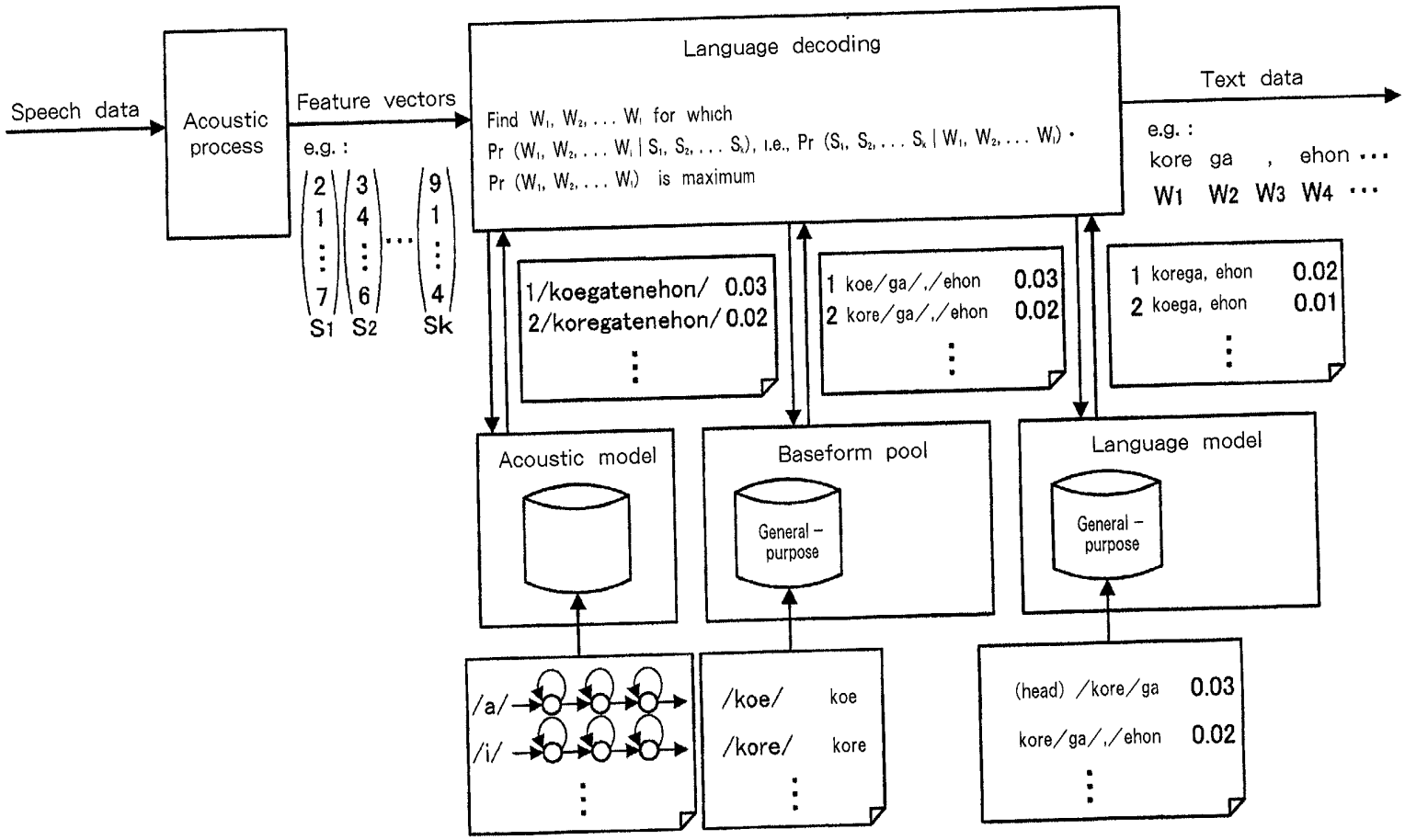

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wl for which
Pr (W1, W2, ... Wl | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wl) ·
Pr (W1, W2, ... Wl) is maximum
Text data
e.g. :
kore ga , ehon ...
W1 W2 W3 W4 ...
1/koegatenehon/ 0.03
2/koregatenehon/ 0.02
1 koe/ga/,/ehon 0.03
2 kore/ga/,/ehon 0.02
1 korega, ehon 0.02
2 koega, ehon 0.01
Acoustic model
Baseform pool
Language model
General – purpose
General – purpose
/a/
/i/
/koe/ koe
/kore/ kore
(head) /kore/ga 0.03
kore/ga/,/ehon 0.02

Fig. 10

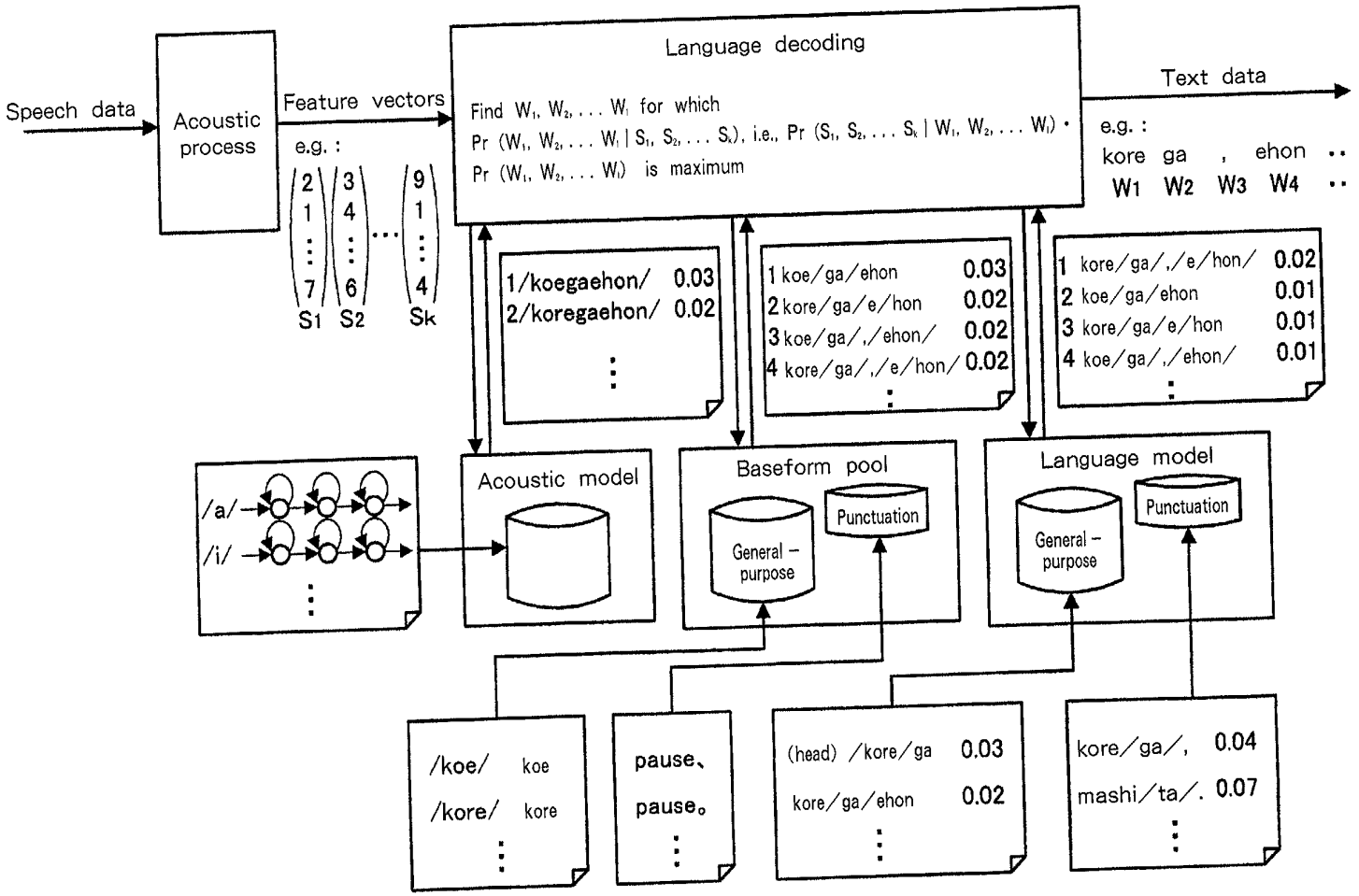

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wi for which
Pr (W1, W2, ... Wi | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wi) ·
Pr (W1, W2, ... Wi) is maximum
Text data
e.g. :
kore ga , ehon ··
W1 W2 W3 W4 ··
1/koegaehon/ 0.03
2/koregaehon/ 0.02
1 koe/ga/ehon 0.03
2 kore/ga/e/hon 0.02
3 koe/ga/,/ehon/ 0.02
4 kore/ga/,/e/hon/ 0.02
1 kore/ga/,/e/hon/ 0.02
2 koe/ga/ehon 0.01
3 kore/ga/e/hon 0.01
4 koe/ga/,/ehon/ 0.01
/a/
/i/
Acoustic model
Baseform pool
General – purpose
Punctuation
Language model
General – purpose
Punctuation
/koe/ koe
/kore/ kore
pause、
pause。
(head) /kore/ga 0.03
kore/ga/ehon 0.02
kore/ga/, 0.04
mashi/ta/. 0.07

Fig. 11

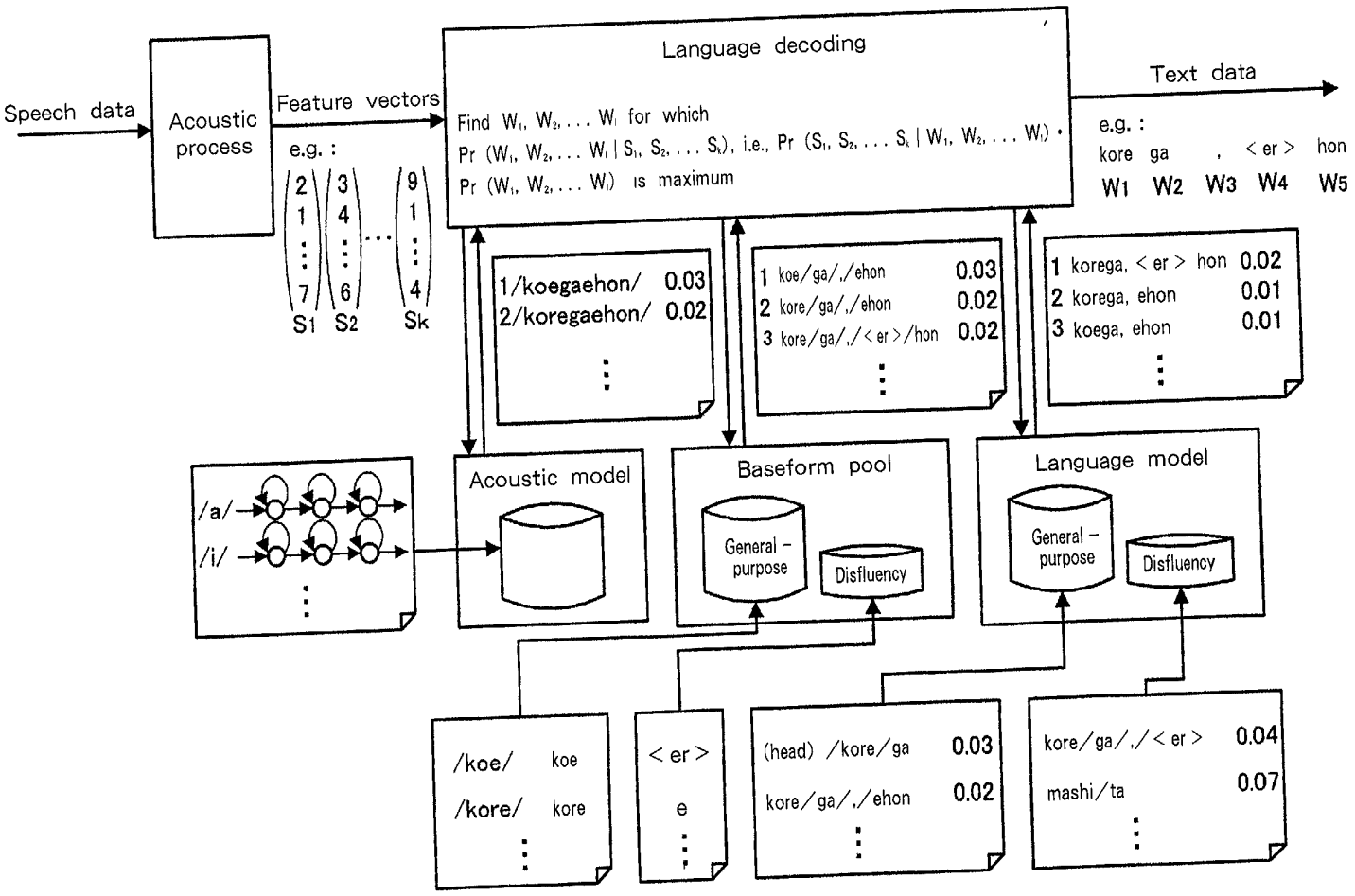

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wi for which
Pr (W1, W2, ... Wi | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wi) •
Pr (W1, W2, ... Wi) is maximum
Text data
e.g. :
kore ga , < er > hon
W1 W2 W3 W4 W5
1/koegaehon/ 0.03
2/koregaehon/ 0.02
1 koe/ga/,/ehon 0.03
2 kore/ga/,/ehon 0.02
3 kore/ga/,/< er >/hon 0.02
1 korega, < er > hon 0.02
2 korega, ehon 0.01
3 koega, ehon 0.01
/a/
/i/
Acoustic model
Baseform pool
General – purpose
Disfluency
Language model
General – purpose
Disfluency
/koe/ koe
/kore/ kore
< er >
e
(head) /kore/ga 0.03
kore/ga/,/ehon 0.02
kore/ga/,/< er > 0.04
mashi/ta 0.07

Fig. 12

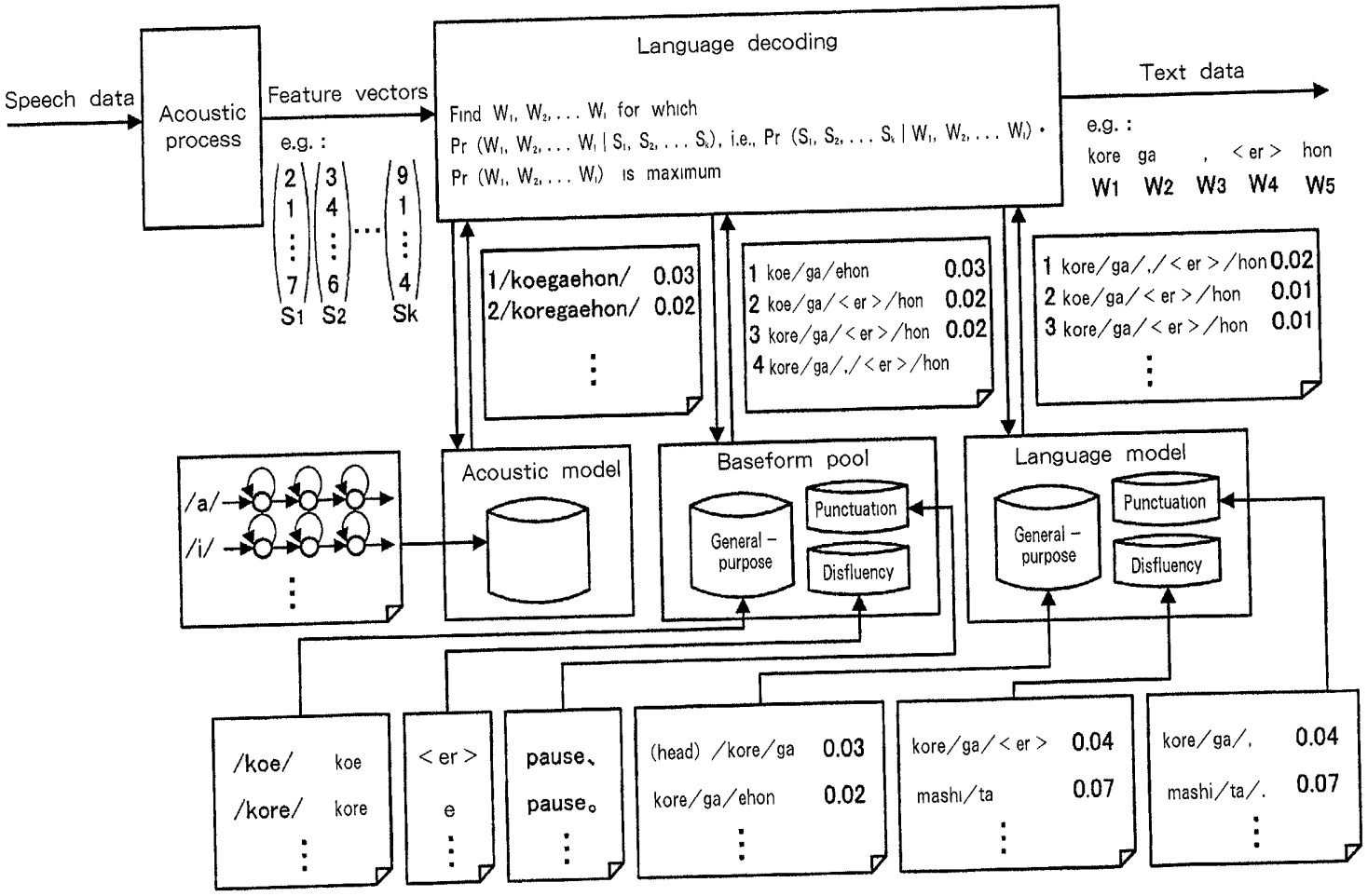

Speech data
Acoustic process
Feature vectors
e.g. :
S1 S2 Sk
Language decoding
Find W1, W2, ... Wi for which
Pr (W1, W2, ... Wi | S1, S2, ... Sk), i.e., Pr (S1, S2, ... Sk | W1, W2, ... Wi) •
Pr (W1, W2, ... Wi) is maximum
Text data
e.g. :
kore ga , < er > hon
W1 W2 W3 W4 W5
1/koegaehon/ 0.03
2/koregaehon/ 0.02
1 koe/ga/ehon 0.03
2 koe/ga/< er >/hon 0.02
3 kore/ga/< er >/hon 0.02
4 kore/ga/,/< er >/hon
1 kore/ga/,/< er >/hon 0.02
2 koe/ga/< er >/hon 0.01
3 kore/ga/< er >/hon 0.01
/a/
/i/
Acoustic model
Baseform pool
General – purpose
Punctuation
Disfluency
Language model
General – purpose
Punctuation
Disfluency
/koe/ koe
/kore/ kore
< er >
e
pause、
pause。
(head) /kore/ga 0.03
kore/ga/ehon 0.02
kore/ga/< er > 0.04
mashi/ta 0.07
kore/ga/, 0.04
mashi/ta/. 0.07

Fig. 13